Abstract

The coronavirus pandemic that has spread has caused various impacts on people's life. The prolonged continuation of the pandemic and the social distancing policies implemented by the government make the public communicate and express their opinions through social media. Twitter is one of the social media with a massive increase in users during the pandemic. The circulation of opinions and information on Twitter can create polemics in the community, thus causing misunderstanding in the community. In the end, this will generate various types of public emotions that can influencing the policies taken by the next government. To be able to find out the public's emotions, emotion classification was carried out on the Indonesian text Twitter related to covid using random forest with TF-IDF and Word2vec feature extraction. Random forest is an ensemble algorithm with supervised learning that is theoretically superior in handling imbalance in data and fast in classifying [1]. The data used through the preprocessing stage, namely case folding, punctuation removal, stopword removal, stemming and tokenization. In this study, the emotional classification of the Twitter text regarding Covid-19 in Indonesian was successfully carried out with a performance value of 0.43 using TF-IDF and 0.38 using Word2vec. As a comparison, data from previous studies [2] were also used which obtained performance value of 0.57 with TF-IDF and 0.60 with Word2vec.

Keywords: random forest, tf-idf, word2vec, emotion classification, covid-19